

PG-Antrag

Wissensmanagement

Lehrstuhl für Künstliche Intelligenz

28. Mai 2001

1. **PG-Thema:** Wissensmanagement
2. **PG-Zeitraum:** WS 2001/2002 und SS 2002
3. **PG-Umfang:** jeweils 8 SWS
4. **PG-Veranstalter:**
 - Katharina Morik, Lehrstuhl Informatik VIII, morik@ls8, GB IV/R115, Tel. 5101
 - Stefan Hausteин, Lehrstuhl Informatik VIII, hausteин@ls8, GB IV/R120, Tel. 2499
5. **PG-Aufgabe:**

Wissensmanagement ist die Sammlung, Zusammenstellung, Präsentation und gezielte Verteilung von Wissen mit dem Ziel, Arbeitsabläufe zu verbessern. Die Arbeitsabläufe können ganz unterschiedlicher Art sein. Wir können an einen Bioinformatiker denken, der Wissen über einen Stoffwechselprozess braucht und dazu verschiedene Datenbanken und Textsammlungen durchsuchen muss. Wir können an eine Marketingabteilung denken, die gezielt Werbematerial versenden möchte und dazu Wissen über Personengruppen (nach Regionen, Alter oder anderen Charakteristiken) braucht. Natürlich sind auch technische Arbeitsabläufe durch Erfahrungen anderer Gruppen zu verbessern. Es verwundert also nicht, dass das Wissensmanagement in letzter Zeit viel Interesse gefunden hat. Smith und Farquar zitieren eine Studie, derzufolge 80% der größten Firmen im Jahre 2000 mindestens ein Projekt zum Wissensmanagement durchführten [4]. Die Nachfrage nach in diesem Bereich ausgebildeten Menschen ist besonders hoch.

In der Projektgruppe sollen die Techniken zum Wissensmanagement studiert werden. Wissensmanagement bedeutet technisch die Integration von

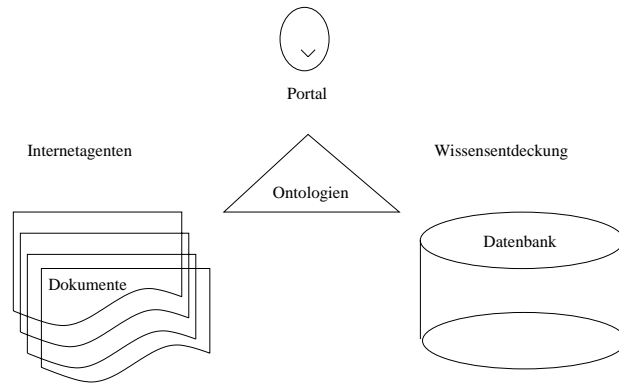


Abbildung 1: Szenario Wissensmanagement

verschiedenen Gebieten: Software-Agenten, Internetbasierte Informationssysteme, intelligente Suchmaschinen, Wissensentdeckung in Datenbanken. Für die Integration spielen Metadaten oder Ontologien eine besonders wichtige Rolle. (vgl. Bild 1).

Strukturierung von Informationssammlungen mit Ontologien Ontologien werden oft als “Strukturierung einer Konzeptualisierung” [2] beschrieben. Dabei werden nicht nur die Konzepte eines Sachbereiches modelliert, sondern auch deren Beziehungen untereinander. Im Wissensmanagement spielen Ontologien eine zentrale Rolle, entweder um das Wissen direkt zu strukturieren, oder bei Dokumentensammlungen als Strukturierung für das Metawissen über den Inhalt der Dokumente. Die Ontologie soll dabei als Klammer für einen einheitlichen Zugriff auf Dokumentensammlungen im WWW und verfügbare Datenbanken dienen.

Dokumentensammlungen im WWW Das WWW kann man als eine einzige grosse Dokumentensammlung auffassen, aber es gibt auch einige zentralisierte und/oder spezialisierte Dokumentensammlungen, die im Internet verfügbar sind. Ein wichtiger Aspekt des Wissensmanagement ist, auf diese Ressourcen einen besseren Zugriff als einfache Suchmaschinen zu gewährleisten. So kann etwa eine Suchanfrage automatisch spezialisiert werden, um irrelevante Dokumente von vornherein auszuschliessen. Oder ein Softwareagent prüft die von der Suchmaschine gelieferten Seiten auf ihre Relevanz und reicht nur eine Auswahl an den Benutzer weiter, oder nimmt zumindest eine Umsortierung vor. Eine weitere Aufbereitungsmöglichkeit ist die automatische Extraktion der betreffenden Passagen in dem gesuchten Text.

Wissensentdeckung in Datenbanken Wissensentdeckung ist das Auffinden interessanter Muster in sehr großen Datenmengen. Das Gebiet beruht auf statistischen Ansätzen, solchen aus der Datenbanktheorie und solchen des maschinellen Lernens. Es gibt inzwischen einige Werkzeuge zur Unterstützung der Wissensentdeckung auf dem Markt. Die Projektgruppe kann das System D-Miner verwenden, das unterschiedliche Verfahren integriert und direkt auf Datenbanken (z.B. Oracle) zugreifen kann.

Ziele und Vorgehen der Projektgruppe Ziel der Projektgruppe ist es, einen Sachbereich durch eine Ontologie-basiertes Wissensmanagementsystem zu strukturieren. Dabei sollen vorhandene Komponenten wie Datenbanken oder Web-Server genutzt und um durch die PG zu erstellende spezialisierte Software ergänzt und insgesamt integriert werden.

Für verschiedene Sachbereiche stehen der Projektgruppe zahlreiche Datenbanken aus realen Anwendungen zur Verfügung, beispielsweise anonymisierte Versicherungsdaten, Abverkaufsdaten einer Ladenkette oder Mobiltelefonaten. Der konkrete Sachbereich für das Portal wird dabei mit der PG abgestimmt. In jedem Fall muss der Sachbereich durch eine geeignete Ontologie strukturiert werden. Dazu sind ein Formalismus und ein entsprechendes Softwaresystem zur Darstellung und Modellierung der Ontologie auszuwählen. Natürlich muss auch die Ontologie selbst modelliert werden. Je nach Sachbereich ergeben sich weitere Anforderungen, die parallel in Arbeitsgruppen bearbeitet werden können. So sind etwa folgende Punkte vorstellbar:

Suchagent: Die gezielte Suche nach Informationen für die Beantwortung komplexer Fragen ist mit Suchmaschinen allein nicht zu bewältigen. Wenn beispielsweise eine Ausbildungsversicherung meist nach der Geburt eines Kindes abgeschlossen wird, so fragt sich, ob es Informationen darüber gibt, wann welche Gruppen von Menschen in verschiedenen Regionen Kinder bekommen. Eine Suchmaschine wie z.B. Google liefert eine Fülle relevanter Seiten, in denen sich auch Antworten befinden, ist aber auf die Sichtung dieser Seiten durch den Menschen angewiesen.

Es könnte also beispielsweise ein Software-Agent entwickelt werden, der aus den Dokumenten der statistischen Bundesämter die wichtigen Seiten für die Beantwortung der Frage nach der Geburtenwahrscheinlichkeit herausucht.

Extraktionsverfahren: Gegeben ein Dokument, das – um im Beispiel zu bleiben – Text und Tabellen über Geburtenraten enthält, finde die relevante Information zur Beantwortung der Fragen, wann Geburten wahrscheinlich sind (bezogen auf das Alter, den Beruf, den Ehestand

der Mutter). Dazu müssen Tabellen und Grafiken extrahiert werden und Textteile aufgefunden werden.

Extraktionsverfahren können für Tabellen, Grafiken und Textteile entwickelt werden. Dabei kann die Relevanz von Wörtern (der Tabellen- bzw. Grafikbeschriftung und in den Textteilen) automatisch ermittelt werden anhand klassifizierter Beispiele.

Wissensentdeckung: Verfahren der Subgruppenentdeckung können Kundengruppen ermitteln. Verfahren zur Behandlung zeitlicher Daten können die Zeiträume beispielsweise zwischen Lebensereignissen (Geburt eines Kindes) und bestimmten Vertragsabschlüssen (Ausbildungsver-sicherung) ermitteln. Dazu müssen die Daten der Datenbank vorverarbeitet werden, damit die Verfahren gute Ergebnisse liefern können. In der Projektgruppe können unterschiedliche Verfahren und Repräsentationen von zeitbezogenen Daten zur Beantwortung komplexer Fragen untersucht werden.

6. PG-Teilnahmevoraussetzungen

Vorbedingung ist mindestens eine Vorlesung aus “Künstliche Intelligenz”, “Maschinelles Lernen”, *oder* “Informationssysteme”. Wünschenswert sind Datenbankkenntnisse oder Kenntnis einer objekt-orientierten Programmiersprache.

7. Minimalziel:

Das Minimalziel der PG ist die Modellierung der Ontologie für den Sachbereich und die Integration von mindestens zwei der oben angegebenen Ansätze in das Wissensmanagement-System.

8. Literatur:

- [1] V. R. Benjamins, D. Fensel, and S. Decker. (KA)²: Building ontologies for the Internet: A mid-term report. *International Journal of Human Computer Studies*, 51(3):687 – 712, 1999.
- [2] Tom R. Gruber. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199–220, 1993.
- [3] S. Robertson and K. Reese. A virtual library for building community and sharing knowledge. *International Journal of Human Computer Studies*, 51(3):663 – 686, 1999.
- [4] Smith and Farquar. The road ahead for knowledge management: An AI perspective. *Artificial Intelligence Magazine*, 21(4):17 – 40, Winter 2000.
- [5] Loren Terveen, Will Hill, and Brian Amento. Collaborative Filtering To Locate, Comprehend, and Organize Collections of Web Sites. *ACM SIGART Bulletin*, 9(3 & 4):10 – 17, 1998.